CERES Data Management Team Working Group Report

October 29, 2019

Katie Dejwakh

CERES Science Team Meeting
Lawrence Berkeley National Laboratory

Outline

CERES DMT Overview

↓

DMT Activity

↓

CERES Code Re-architecture

↓

Systems & CM

DMT Overall Tasks

Algorithms

- Implementation
- Verification
- Validation assistance

$$Z = \sum_{i=0}^{n} (\Delta C_i/\sigma C_i)^2 + \sum_{j=1}^{m} (\Delta v_j/\sigma v_j)^2 + \sum_{k=1}^{l} (\Delta F_k/\sigma F_k)^2 \qquad (3)$$

where on the r.h.s. the first term represents the cloud fraction adjustr
er model variable adjustments, and the third term the flux compon

Equation (4) below restricts the solution such that the sum of the cl
This prevents unrealistic solutions (i.e.., sum of adjusted total fracti

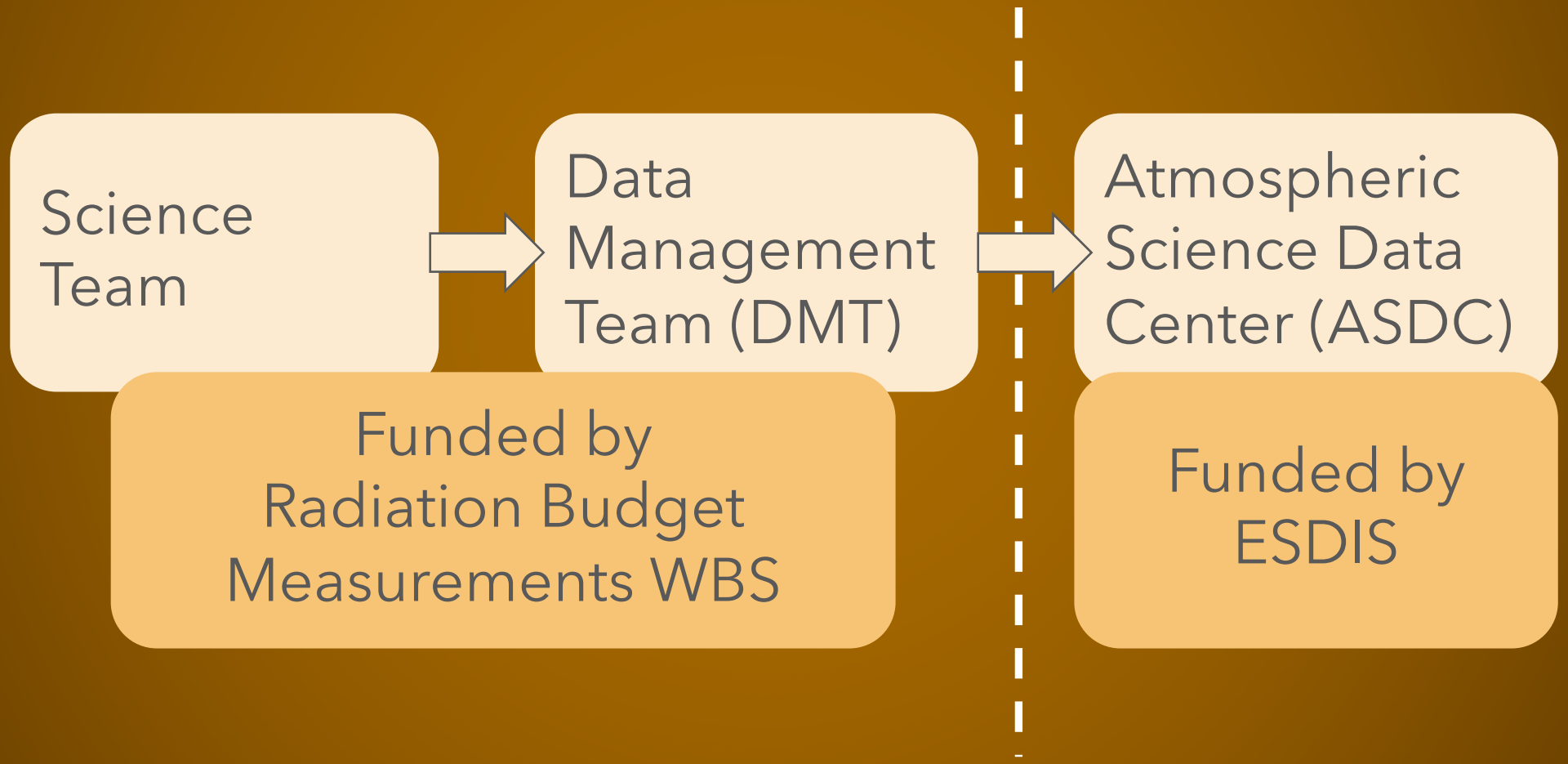
$$X = \sum_{i=0}^{n} \Delta C_i = 0 \qquad (4)$$

Software

- Maintenance
- Configuration management

```
if ( dato(i,1) == -1 .and. dat(i,1) > 0.0 .and. dat(i,1)< 10000 ) then

 if (ifill == 1) then
  qc_validflag_sahclim_global (i) = 1
 endif

 if (ifill == 2) then
  qc_validflag_epoch_global (i) = 1
 endif

 qc_validflag_all_global(i) = 1

 dato(i,1:np) = dat(i,1:np)
endif

if ( dato(i,1) == -1  .and. ifill .eq. 2 ) then !LAST RESORT IGBP Based
 igbp1 = igbp(i)
 u0ohs = 1.0
 wv = 1.0 ! Pw(cm)
 call land_spec (igbp1, u0ohs, wv,  spec18_dum, bbalb)
 dato(i,1) = bbalb*10000
 dato(i,2) = u0ohs *1000
```

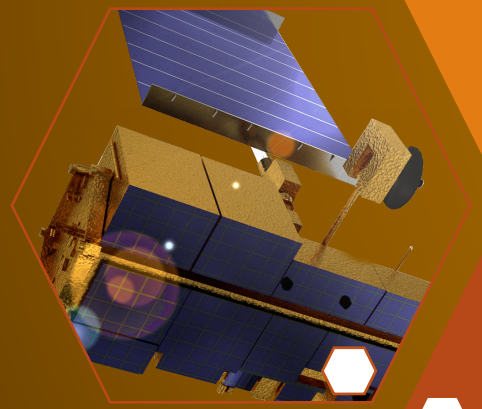
DMT-ASDC Interface



Recent and Ongoing DMT Activity

MODERNIZATION

Production Code

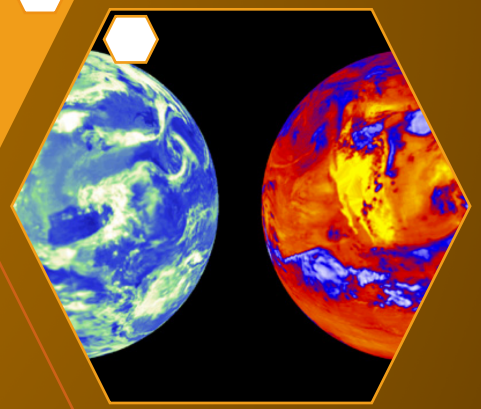


Deprecated file formats

Contingency Planning (workforce, spacecraft)

Million lines of code

User experience (metadata, display)



File Formats

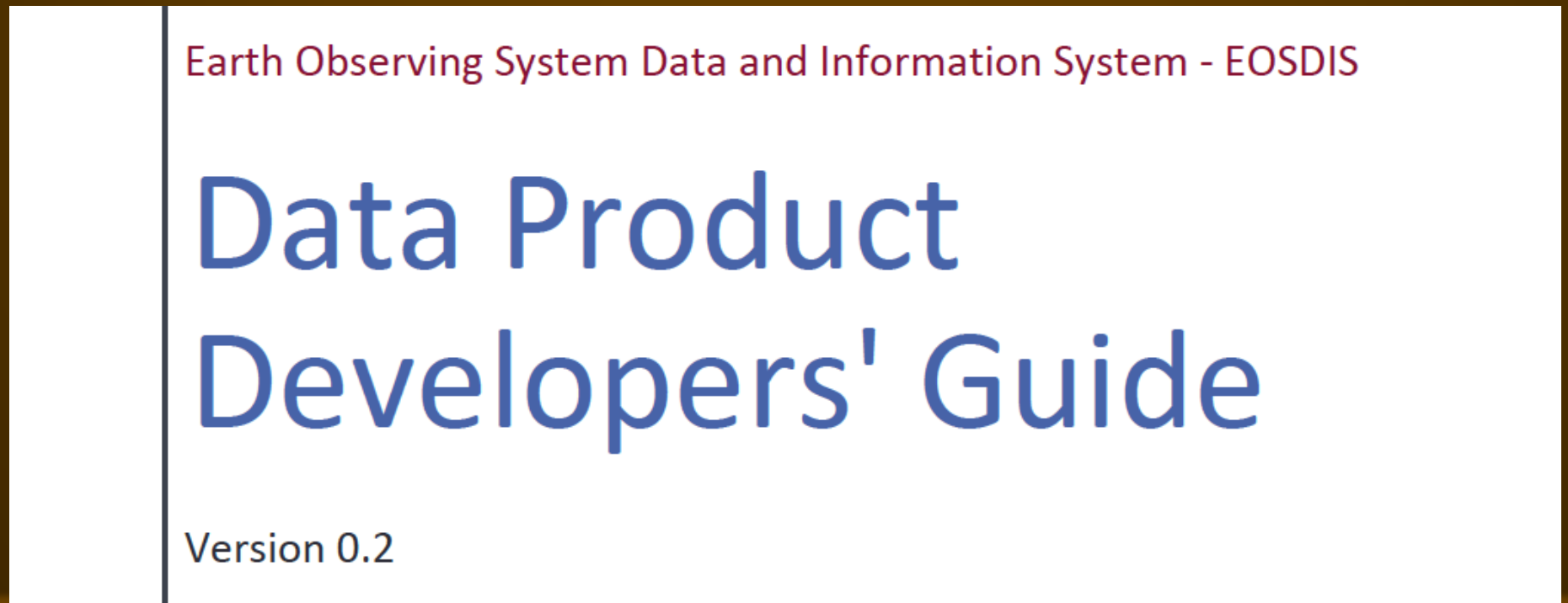
- Eliminating extraneous files:
  - IES
  - Binaries
  - Temp
- Output file format:

HDF4 → HDF5

- File format code deliveries start: mid-December, 2019

Metadata

- Universal Metadata Model
- CF Compliance
- Data Product Developers' Guide (DPDG), Beta


Earth Observing System Data and Information System - EOSDIS

Data Product Developers' Guide

Version 0.2

Subsetter/Ordering Tool

- Emphasis on availability and flexible ordering
- Migrations:
  1. As-is virtualization
  2. Redesigned, containerization (OpenShift)

Energy Balanced and Filled (EBAF)

Climate Data Record (CDR) of monthly TOA fluxes and consistent computed surface fluxes and clouds suitable for analysis of variability at the intra-seasonal, inter-annual, and longer time scales.

Data Product Information	Description	Parameter	Resolution	Version/ Availability	Order Data
EBAF	Monthly and climatological averages of observed TOA and computed surface all-sky, clear-sky (spatially complete), and cloud radiative effect (CRE) fluxes and associated clouds. The TOA net flux is constrained to the ocean heat storage. EBAF Data Quality Summary				Browse & Subset

Synoptic TOA and surface fluxes and clouds (SYN)

Product Availability

Edition 4 Terra and Aqua

Product	Platform	Processed Thru	Publically Available?
BDS	Terra, Aqua	June '19	Yes
SSF		June '19	Yes
SSF1deg-Hour		May '19	Yes
SSF1deg-Day/-Month		May '19	Yes
SYN1deg-Hour/3Hour/MHour	Terra+Aqua	May '19	Yes
SYN1deg-Day/-Month		May '19	Yes

Edition 4 Terra and Aqua

Product	Platform	Processed Thru	Publically Available?
CldTypHist	Terra+Aqua	Apr. '19	Yes
FluxByCldTyp	Terra+Aqua		Fall '19
EBAF ToA+Surface	Terra+Aqua	May '19*	No*
EBAF ToA	Terra+Aqua	June '19	Yes

*Under review for release, post-Science Team Meeting

Edition 1 S-NPP

Product	Platform	Processed Thru	Publically Available?
BDS	Terra, Aqua	July '19	Yes
SSF		July '19	Yes
SSF1deg-Hour		June '19	Yes
SSF1deg-Day/-Month		June '19	Yes
SYN1deg-Hour/3Hour/MHour	Terra+S-NPP	June '19	Yes
SYN1deg-Day/-Month		Nov. '17	Yes

Upcoming/Recent Editions

Ed 1 J1/NOAA-20	• Sensor gain has stabilized • Accommodate LW channel
Ed 2 S-NPP	• Incorporate updated VIIRS data • FM5: same radiometric scale as FM3
EBAF 4.1	• Only either C5 or C6.1 MODIS • C6.1-based MATCH aerosols • Corrected Clouds inconsistencies

Code Re-architecture

CERES Code Goals

First Priority

- Readability
- Maintainability
- Modularity
- Documentation

Second Priority

- Modernization
- Run-time Efficiency

TISA

- Baseline: Edition 4 code
- Short-term: low-hanging fruit
- Long-term: Infrastructure changes – C++
- Weekly code review meetings
- Use of:
  - Jira/Bitbucket/Confluence
  - “Understand” static code analysis tool
  - cpd (Copy-paste Detector)

TISA

LOW-LEVEL (Nov. ’18)

- Remove “dead” code
- De-duplicate
- Generalize
- Collect multi-purpose routines for library

HIGH-LEVEL (Apr. '19)

- Designing future system
- Build general library
- Spatiotemporal flexibility

TISA Code Reduction

Product	LoC Before	LoC After	% Reduction
SSF1deg	1,598	1,433	10.3%
SSF1deg-Hourly	11,835	11,718	0.9%
SSF1deg post-proc.	4,086	3,921	4.0%
SSF1deg-Month	20,275	16,941	16%
TSI	40,945	34,986	15%
GGEO Driver	5,776	1,448	75%
GGEO Daily/Monthly	8,729	5,929	32%
SYN1deg	34,847	31,248	11%

TISA Code Reduction

Product	LoC Before	LoC After	% Reduction
SSF1deg	1,598	1,433	10.3%
SSF1deg-Hourly	11,835	11,718	0.9%
SSF1deg post-proc.	4,086	3,921	4.0%
SSF1deg-Month	20,275	16,941	16%
TSI	40,945	34,986	15%
GGEO Driver	5,776	1,448	75%
GGEO Daily/Monthly	8,729	5,929	32%
SYN1deg	34,847	31,248	11%

Clouds

- Baseline: Edition 4 MODIS/VIIRS & GEO code
- Goals:
  - Clean up code
  - Generalize for enhanced library
  - Improve validation capabilities
  - Improve code documentation

Clouds

LOW-LEVEL (Nov. '19)

- Remove "dead" code
- Remove unreachable code
- Decrease use of "magic numbers" (hard-coded values)

HIGH-LEVEL (Winter '20)

- Collect core, common functionality into library
- Robust use of Git
- Improve validation strategy

Systems & CM

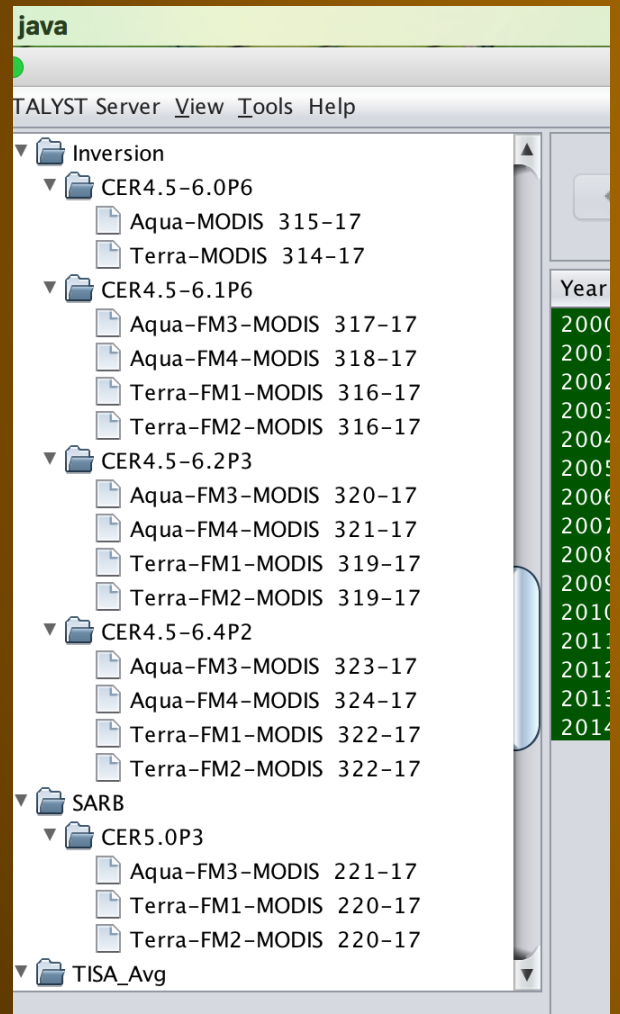
Production

- Production can be closed:
  - Production requests (PRs) in PR Tool
  - In-console CATALYST system
- PR Tool runs in OpenShift
- ALL PGEs RUN ON x86 MACHINES

Hardware Updates

- New x86 blades installed
  - Overall: 55% wall-time reduction
  - Biggest gain:
    - SYNI processing
    - 78% wall-time reduction
- Subsequent computers → virtualized
- CERES websites: migration in progress

Software Updates



- Accommodating SFTP
- Preliminary Instrument code delivery for Edition 2 NPP
- CERES websites: migration to new environments in progress
  - New CERES website design – Spring 2020
  - Working group pages - consolidation

Cloud-Compatibility

- PaaS: OpenShift
- Docker containers
- Production proof-of-concept:
  - TISA PGEs can run: 7.3.1P2, P3, and P4 (TSI)
  - On pace with MAIA production development
  - “Vertical slicing” next (chain)

ASDC Long-term Plans

- CERES as a SIPS
- De-coupling of archive and PGE data access/storage
- Pros:
  - More of a technical issue – we run production now
  - Ingest has been brittle – we can direct improved infrastructural design
- Cons:
  - Extricating responsibility/$ will take time
  - CERES itself will need more SA skill

Conclusions

- Technical capacity-building
  - Cloud/container technology
  - Software development
  - IT security compliance
- Production running well
- Preparing for new editions and data formats